AN ABSTRACT OF THE THESIS OF

Haiou Wang for the degree of Master of Science in Computer Science

presented on June 21, 2001.

Title: Logic Sampling, Likelihood Weighting and AIS-BN: An Exploration of Importance Sampling

Abstract approved: Redacted for privacy

Bruce D. D'Ambrosio

Logic Sampling, Likelihood Weighting and AIS-BN are three variants of stochastic sampling, one class of approximate inference for Bayesian networks. We summarize the ideas underlying each algorithm and the relationship among them. The results from a set of empirical experiments comparing Logic Sampling, Likelihood Weighting and AIS-BN are presented. We also test the impact of each of the proposed heuristics and learning method separately and in combination in order to give a deeper look into AIS-BN, and see how the heuristics and learning method contribute to the power of the algorithm.

Key words: belief network, probability inference, Logic Sampling, Likelihood Weighting, Importance Sampling, Adaptive Importance Sampling Algorithm for Evidential Reasoning in Large Bayesian Networks(AIS-BN), Mean Percentage Error (MPE), Mean Square Error (MSE), Convergence Rate, heuristic, learning method.

Logic Sampling, Likelihood Weighting and AIS-BN: An Exploration of Importance Sampling

by

Haiou Wang

A Thesis

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Completed June 21, 2001
Commencement June 2002

Master of Science thesis of Haiou Wang presented on June 21, 2001

APPROVED:

Redacted for privacy

Major Professor, representing Computer Science

Redacted for privacy

Chair of the Department of Computer Science

Redacted for privacy

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Redacted for privacy

Haiou Wang, Author

ACKNOWLEDGMENTS

I would like to thank all those people who helped me during my study and research, Dr. Bruce D'Ambrosio, Dr. Tadepalli Prasad and the anonymous referees.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure | Page

LOGIC SAMPLING, LIKELIHOOD WEIGHTING AND AIS-BN: AN EXPLORATION OF IMPORTANCE SAMPLING

CHAPTER 1
INTRODUCTION

Bayesian networks are getting more and more popular as a modeling tool for complex problems involving reasoning under uncertainty. Since it is impossible to perform inference by exact methods in very large Bayesian networks, approximate inference seems to be the only computationally feasible alternative. There are two basic classes of approximate schemes: stochastic sampling and search based algorithms. We only focus on the former class, stochastic sampling algorithm, in our research here.

Stochastic sampling algorithm is also called Monte Carlo sampling, stochastic simulation or random sampling. The basic idea of stochastic sampling is to estimate the frequency of the interested event in a set of samples instead of estimating the probability of the event directly. In other words, the stochastic simulation methods use the network to generate a large number of concrete samples of the domain that are consistent with the network distribution. These methods give an approximation of the exact evaluation.

The precision obtained by stochastic sampling generally increases with the number of samples generated and is slightly affected by the network size. The execution time is almost independent of the topology of the network and is linear in the number of the samples.

In the thesis, I address the following three algorithms, which are related yet also different.

- Logic sampling.
- Likelihood Weighting
- AIS-BN: Adaptive Importance Sampling for Evidential Reasoning in Large Bayesian Networks.

The remainder of the thesis is presented in four parts. Part 2 describes and compares the algorithms. Part 3 shows some impressive testing results on the alarm network, which supports the published claims for AIS-BN. Part 4 describes the empirical experiments on the Computed-based Patient Case Study(CPCS) network with the three algorithms and also several versions of AIS-BN. Part 5 summarize the results and suggests several possible studies that we can do in the future.

CHAPTER 2
ALGORITHMS FOR LOGIC SAMPLING, LIKELIHOOD WEIGHTING AND AIS-BN

This part will outline the three algorithms in our research.

In the following discussion, X denotes to the set of nodes in Bayesian network. x_i denotes a specific node i in Bayesian network. E denotes the set of evidence nodes. X-E denotes the set of the difference between X and E. Pa(x_i) denotes the set of parents of node x_i. P denotes to Probability. Pr denotes the joint probability. Pr(X) is the joint Probability over the set X in the Bayesian network, and Pr(X)-old denotes that in the old Bayesian network while Pr(X)-new denotes that in the new Bayesian network.

We know the joint probability is the product of the probability distributions over each of the nodes conditional on their pqrents in the Bayesian network model, i.e.,

$Pr(X) = \prod_{i=1}^{n} P(x_i/Pa(x_i))$.

2.1 Logic Sampling

Logic Sampling is the simplest and the first proposed sampling algorithm for Bayesian networks (Henrion, 1986). See figure 2.1.

It repeats the simulations of the world described by the Bayesian network, and estimates probability according to the frequencies with which relevant events occur. Each round of the simulation starts by randomly choosing a value for each root node of the network, weighing the choice by the prior probability. It then proceeds to sample the immediate children conditioned on the sample values selected for the parents. In order to estimate $P(X/E)$, we repeat the process many times and then compute the ratio of the number of runs where X and E are both true to the number of runs where only E is true. The prior probability of the root nodes and the probability conditioned on parents that we use in sampling process are also known as importance function.

```
1.Order the nodes according to their topologic order.
2.Initialize the importance function according to the CPT from
  the original Bayesian network.
3.For I = 1 to m do
     3.1 Generate a sample based on the importance function;
     3.2 If evidence is true
         Increment N (E) by 1.
         If X is also true increment M (X, E) by 1.
4.P (X/E) = M (X, E) / N (E)
```

FIGURE 2.1: Logic Sampling Algorithm to calculate Pr(X/E)

The algorithm will always converge to the desired solution, although it might take many runs. The main problem is: when the assignment of values to E rarely happens, we have to discard lots of samples, because it is so difficult to get a useful sample satisfying the assignment to E. This problem is gotten around by the approach of Likelihood Weighting.

2.2 Likelihood Weighting

Figure 2.2 gives the basic framework for Likelihood Weighting (Fung & Chang, 1989, Shachter & peot, 1989). The basic idea of Likelihood Weighting is almost the same as Logic Sampling. But in Likelihood Weighting, whenever we reach an evidence variable, instead of randomly choosing a value for it, we take the given value. We then use the prior conditional probability of the evidence value as a 'weight' for the sample in counting the ratio.

```
1.Order the nodes according to their topologic order.
2.Initialize the importance function according to the CPT from
  the original Bayesian network.
3.For I = 1 to m do
3.1 Generate a sample based on the importance function.
            If the variable is not a part of the evidence. Pick
            according to the CPT just like what  we do in LS.
            If the variable is part of the evidence, set it to
            the value indicated by the evidence.
     3.2 Calculate the weight as:
            Weight = Pr(E).
     3.3 N (E) = N(E) +weight
     3.4 If X is true, add weight to M (X, E).
4. P (X/E) = M (X, E) / N (E)
```

FIGURE 2.2: Likelihood Weighting Algorithm to calculate Pr(X/E)

2.3 AIS-BN

AIS-BN(Cheng and Druzdzel, 1999), Adaptive Importance Sampling for evidential reasoning in large Bayesian networks purports to be a more advanced version of the above two. It introduced the following new ideas:

(1)In the process of simulation, a smooth learning method is used to adapt conditional probability table(CPT). In LS and LW, the CPT stays the same as the original belief network throught out the whole process.

(2)Before the simulation starts, two heuristics are used to make the process converge more smoothly.

We know that the failure of most stochastic sampling algorithms is due to the failure of generating useful samples, especially in a large belief network with extremely small probabilities in the CPTs. AIS-BN generates the samples based on a new network which has the same structure but different CPTs as the original one, which is the result of using the heuristics and the learning method. It is more likely to get useful samples using the adapted network.

Then, what is Importance Sampling? Importance Sampling (Shachter & Peot, 1989) is different from LS and LW. In Importance Sampling, The importance function will be updated using the samples it has generated. And the new samples will always be generated according to the newest importance function. The generic Importance Sampling algorithm is described in figure 2.3. Refer to the paper AIS-BN: An Adaptive Importance Sampling Algorithm for Evidential Reasoning in Large Bayesian Networks (Cheng and Druzdzel, 1999) for the math foundations.

In step 3.1 above, we can have two different variants. One is to generate samples as is done in LS, and the other as is done in LW. Step 3.5 also has different variants, the purpose of updating the importance function is to get a better convergence rate. In AIS-BN, we use a smooth learning method to update importance function based on the samples generated in the specific interval. Figure 2.4 is the learning method for AIS-BN.

```
1. Order the nodes according to their topologic order.
2. Initialize the importance function  according to the CPT
   from the original Bayesian network.
3. For I = 1 to m
      3.1 Generate a sample based on the importance function.
      3.2 Calculate the weight as:
          Weight =Pr(X)-old/pr(X-E)-new
      3.3 N (E) = N(E) + weight
      3.4 If X is true, add weight to M (X, E).
      3.5 if (mod I  updating-interval) = 0
             update the importance function
4. P (X/E) = M (X, E) / N (E)
```

FIGURE 2.3: Generic Importance Sampling Algorithm to calculate Pr(X/E)

Input: current importance function P_k, learning rate η
Output: new importance function P_{k+1}
Based on the samples generated in the current learning interval, use the following formulation to update the ICPT. Return the ICPT to the caller.
$P_{k+1} = P_k + \eta(P'_k - P_k)$, P'_k is estimated by the samples generated in updating interval K.

FIGURE 2.4: Learning Method in AIS-BN

AIS-BN is an adaptive version of the Importance Sampling. The main algorithm is stated in figure 2.5. Before step 3, the real Importance sampling process, two heuristics are called. CPT in our paper refers to the original conditional probability table. ICPT refers to the table that has been updated.

1. Heuristic 1, which initializes the ICPT tables of the parents of evidence to uniform distributions.
2. Heuristic2, which updates the value in the ICPT according to the threshold θ. Any probability in the network less than θ will be replaced by θ, at the same time, the largest probability p in the same ICPT will be subtracted by $(\theta - p)$.
3. Importance Sampling algorithm (See figure 2.3)

FIGURE 2.5: AIS-BN Algorithm

There are many variants for Importance Sampling, and people can put extra weight coefficient to the samples, which might also improve the performance of the algorithm.

CHAPTER3
SOME EXPERIMENTS ON THE TYPICAL ALARM NETWORK

3.1 Experiments Description

To demonstrate the power of the AIS-BN, we executed two sets of experiments on the following alarm network (Artificial Intelligence A modern Approach, Russell, S., and Norvig, R., 1995). As mentioned in part 2, in the Importance Sampling, there are two variants in generating samples. The first set of experiments is designed for the one generating samples as is done Logic Sampling, and the second set is for the one as is done in Likelihood Weighting. We name them AIS-BN-LS and AIS-BN-LW, respectively.

As mentioned in part 2.3, the idea of AIS-BN is to sample on a new network with the same structure yet different CPTS from the original belief network, in which it will be more likely to obtain useful samples. We performed empirical tests comparing the AIS-BN algorithm to the Logic Sampling and Likelihood Weighting to show that the idea is practicable. Furthermore, in order to make a deeper analysis of the two heuristics and the smooth learning method of the AIS-BN, we divided AIS-BN into four versions:

(1) Basic AIS-BN, Importance Sampling with the smooth learning method, denoted by AIS-BN0.

(2) Basic AIS-BN plus heuristics1, which initializes the ICPT tables of the

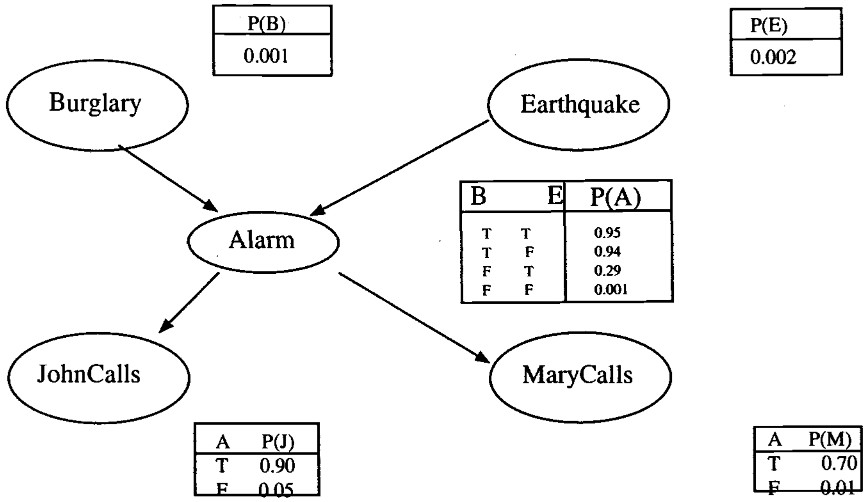

FIGURE 3.1: Alarm Network

parents of evidence to uniform distributions, denoted by AIS-BN1.

(3) Basic AIS-BN plus heuristics2, which increases the samll values in the ICPT according to the threshold θ, denoted by AIS-BN2

(4) AIS-BN, which includes the smooth learning method and both heuristics.

Experiments for each version of AIS-BN were executed in order to look more deeply into the AIS-BN so that we can get an idea how the heuristics and the smooth learning method contribute to the power of the algorithm.

Every set of experiment consists of 5 queries. Each query belongs to one of the following four categories of inference in the Bayesian networks. The queries were picked randomly, and so were the evidence nodes and their value for each query. The number of evidence nodes in each query ranges from 1 to 3.

(1) Diagnostic inference(from effects to causes). That is we have evidence on effects and wish to capture posterior probability of the causes. For example, Given that JohnCalls, infer P(Earthquake/JohnCalls).

(2) Causal inference (from causes to effects). We know the evidence of causes

and wish to know the conditional probability of the effects. For example, infer P(JohnCalls/Burglary given Burglary is *fales*.

(3) InterCausal inference, which is between causes of a common effect. We know the evidence of some causes and wish to know the conditional probability of other causes. All cause nodes in this case have the same descendent. For example, infer P(Burglary/Earthquake) given Earthquake is *true*.

(4) Mixed inference, a combination of two or more of the above.

We measured convergence rate by Mean Percentage Error(Error), i.e. the difference between test value and real value expressed as a percentage of the real value.The more powerful the algorithm, the lower the MPE should be.

$MPE = \frac{1}{n}\sum_{i=1}^{n} 100 * |testvalue_i - realvalue_i| / realvalue_i$

Since there are several tunable parameters used in AIS-BN, we list the parameters we used in our test on the Alarm Network: learning rate 0.2, updating interval 100, threshold 0.1.

3.2 Experiments Result

Figure 3.2 and Figure 3.3 are two plots from the first experiment set. Figure 3.4 and Figure 3.5 are from the second experiment set.

The superiority of AIS-BN-LS to LS and LW is not obvious from Figure 3.1, but AIS-BN-LS is comparable to the traditional Likelihood Weighting. We find a very funny phenomenon in Figure 3.2: The AIS-BN-LS is worse than AIS-BN-LS2, which means AIS-BN0 with heuristic 1 alone works better than

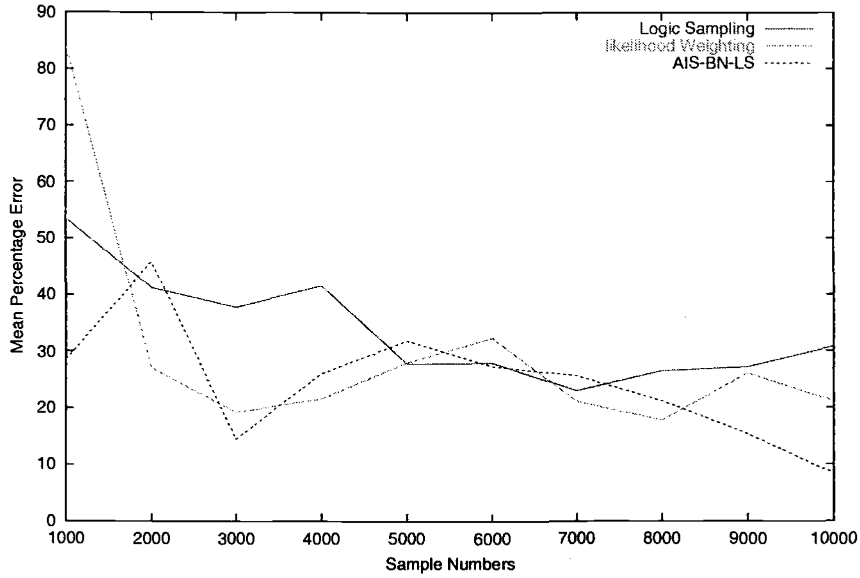

FIGURE 3.2: Plot of convergence of LS, LW and AIS-BN-LS

AIS-BN0 with both heuristics.

Figure 3.3 presents the dramatic improvement of using the heuristics and smooth learning method to the traditional Likelihood Weighting.

Both Figure 3.3 and Figure 3.5 shows the following information: Heuristic 2 (See Part 2) seems to be more critical. AIS-BN-LS2 performs the best in Figure 3.3. AIS-BN-LW2 is almost as good as AIS-BN-LW in Figure 3.5. The result can be very sensitive to the threshold θ used in Heuristic 2, which is 0.1 in both sets of experiments here.

Comparing Figure 3.2 and Figure 3.4, AIS-BN-LW works a way better then AIS-BN-LS. We know the performance of Likelihood Weighting is better than

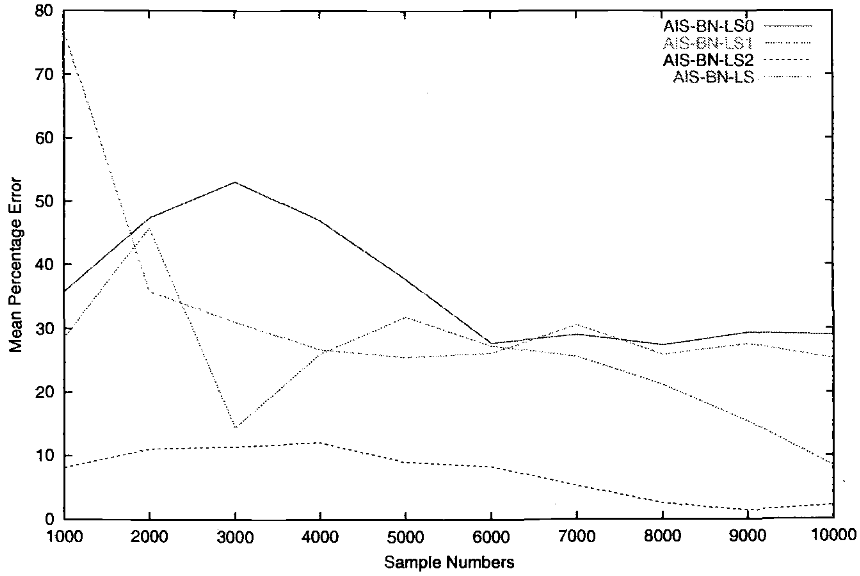

FIGURE 3.3: Plot of convergence of the four versions of AIS-BN-LS

Logic Sampling, which is a nice explanation of why AIS-BN-LW seems more advanced.

Now, We know the new ideas AIS-BN introduced work pretty well in the Alarm network. How will they perform in a large belief network? We address this issue in next part.

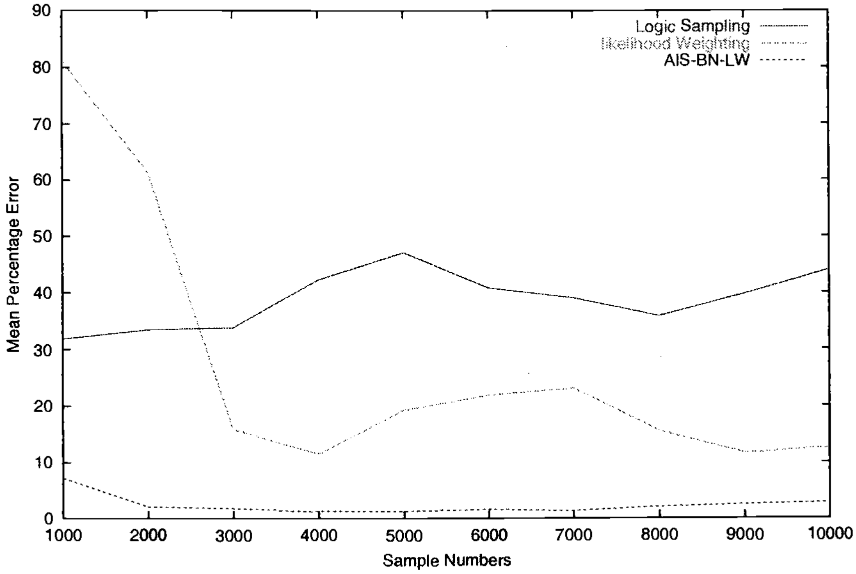

FIGURE 3.4: Plot of convergence of LS, LW and AIS-BN-LW

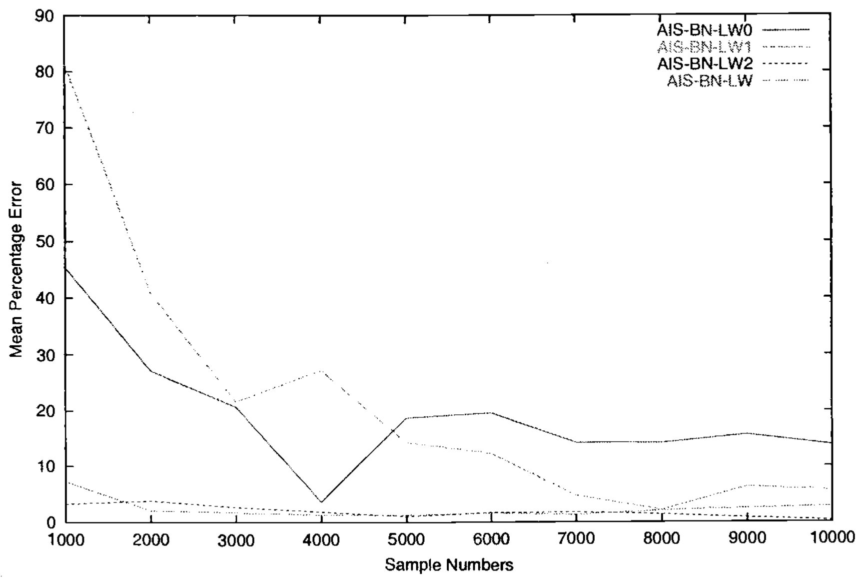

FIGURE 3.5: Plot of convergence of the four versions of AIS-BN-LW

CHAPTER 4
EXPERIMENTS ON CPCS

4.1 Experiment Description

AIS-BN is designed for evidential reasoning in Large Bayesian Networks. The experiments and results we present in this part were performed on Computed-based Patient Case Study network (CPCS) (Pradhan et al., 1994), one of the largest and most realistic networks available to the scientific community.

CPCS network is a large multiply connected multi-layer network consisting of 422 multi-valued original nodes and 1043 noisy max nodes, and covers a subset of the domain of internal medicine. Among the 422 nodes, 14 nodes describe diseases, 33 nodes describe history and risk factors, and the remaining 375 nodes describe various findings related to the diseases. In our research, we used parent divorcing and Temporal Transformation to represent the multiplicative factorization of noisy-max.

The experiment we executed on CPCS is very similar to what we did on the Alarm network. The variant we used in this part is AIS-BN-LW. For convenience, we just call it AIS-BN.

We generated 20 testing queries in our experiment. Each query and its value were picked randomly from the original 422 nodes in CPCS. The evidence nodes for each query were also picked randomly from the 422 nodes. The number of evidence nodes in the query ranges from 10 to 25, and the probability of the evidence ranges from 10^{-12} to 10^{-38}. Each case will be executed 5 times by LS, LW, AIS-BN0, AIS-BN1, AIS-BN2 and AIS-BN.

The parameters used in AIS-BN are listed below:

threshold $\theta = 0.04$,

updading-interval = 1000.

We use the following formula to calculate the learning rate η ,which is proposed by Cheng and Druzdzel in their paper. η decais in the learning Process.

$\eta(k) = 0.4 * (0.35)^{k/10}$

In our AIS-BN, we only update the AIS-BN in the first 10 leaning iteration.

Mean Square Error (MSE), i.e., the sum of square differences between the real value and the test value, was used to measure the convergence rate in this part.

$MSE = \frac{1}{n} \sum_{i=1}^{n} (testvalue_i - realvalue_i)^2$

4.2 Experiment Result

Figure 4.1 is the plot showing the convergence rate in LS, LW and AIS-BN. Figure 4.2 illustrates the convergence rate of AIS-BN in different versions.

AIS-BN shows great superiority in Figure 4.1. In the test, we find the algorithm converges after a couple of hundred samples.

Comparing the different versions of AIS-BN from Figure 4.2, uniforming the distribution of parents nodes of the evidence (Heuristic 1) is more importance

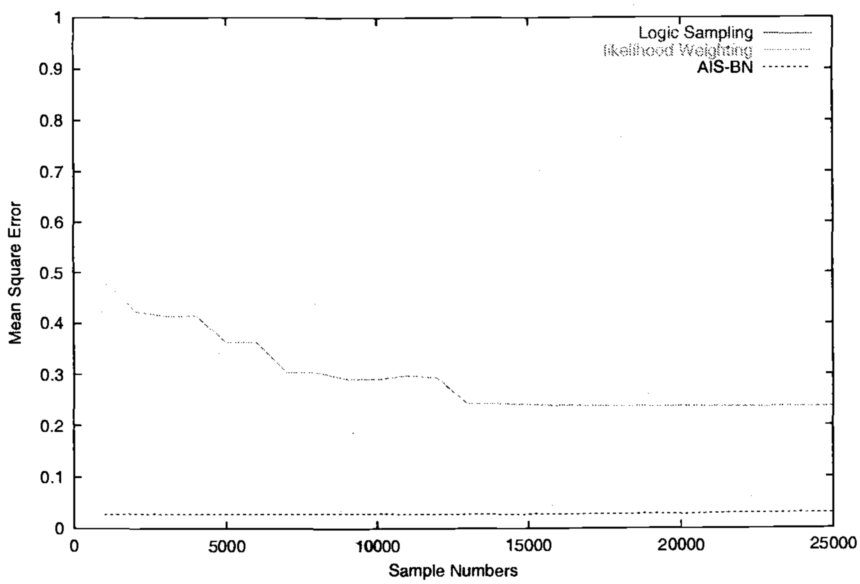

FIGURE 4.1: Plot of convergence of LS, LW and AIS-BN on CPCS

than updaing the CPTS According to the Threshold (Heuristic 2). Although neither of the heuristics works perfectly by itself, the combination of the heuristics and smooth learning method turns out to have huge chemistry.

Comparing Figure 3.2 and Figure 3.3, Figure 3.4 and Figure 3.5, and Figure 4.1 and Figure 4.2, we find Importance Sampling with learning is about same as either Likelihood Weighting or Logic Sampling, depending on how it samples. The learning method seems to make the result more stable after generating a certain number of samples.

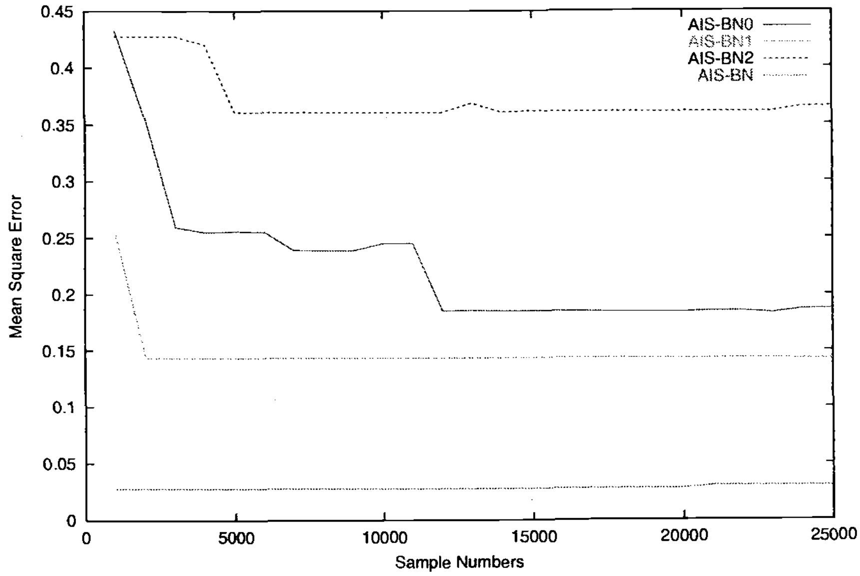

FIGURE 4.2: Plot of convergence of the four versions of AIS-BN on CPCS

CHAPTER5
DISCUSSION AND CONCLUSION

In the research, we studied Importance Sampling in Bayesian networks. Both Chapter 3 and Chapter 4 support that using AIS-BN leads to more rapid convergence in large Bayesian networks than we found with Logic Sampling or Likelihood Weighting. Both heuristics and the learning method play certain roles in AIS-BN algorithm. Neither of them seems dominant, but together they show a major contribution to the algorithm.

Also in AIS-BN, the time spent on learning the importance function trades off obviously the time spent on sampling. The experiment above didn't take into account this trade-off. It might be more reasonable to stop learning at some point when the importance function is good enough. Actually we only let our code learn during the first 10 learning intervals.

Performance of AIS-BN relates very closely to those tunable parameters in the algorithm, such as updating interval and threshold in the second heuristics. Different networks seem prefer different parameters. We use 0.1 and 0.04 respectively for the test on Alarm network and CPCS network in our research. Study on how to choose the parameters for different networks is an interesting future work.

In the current AIS-BN, we initialize the importance conditional probability table (ICPT) of every parent node of the evidence E to uniform distribution. Since we find in chapter 3 that heuristic 2 doesn't work so well with the Alarm network, we definitely can do some further work with it. We can pre-compute the prior probability distribution, $Pr(E)$ for the evidence nodes, $Pr(E)$. Only when $Pr(E)$ is lower than some value, will the ICPT be modified to the uniform distribution.

When learning, we can view the process as a network rebuilding process. The AIS-BN constructs a new network whose structure is the same as the original one with different condition probability tables (CPTs). Current algorithm uses a smooth learning method. Some other learning methods, such as adjusting the learning rate according to the error, might also be applicable and even better.

Additionally, statistics based on the categories (See Part 3) of the query given to the algorithms is also a tempting future work.

BIBLIOGRAPHY

[1] Jian Cheng a nd Marek j. Druzdzel *AIS-BN: An Adaptive Importance Sampling algorithm for Evidential Reasoning in Large Bayesian Networks* . Journal of Artificial Intelligence Research 1,1999, 1-37

[2] Li and D'Ambrosio *Efficient Inference in Bayes Nets as a Combinatorial Optimization Problem* . International Journal of Approximate Reasoning, Vol 11 No 1 pp 55-81

[3] D'Ambrosio *Local Expression Languages for Probabilistic Dependence* . International Journal of Approximate Reasoning, Vol 13-1, pp 61-81, July, 1995

[4] Takikawa and D'Ambrosio *Multiplicative Factorization of Noisy-Max* Uncertainty in Artificial Intelligence 5, pp 622-630, 1999

[5] D. Heckerman *Causal Independence for Knowledge Acquisition and Inference* In Processing of the Nineth Conference on Uncertainty in Artificial Intelligence, Pages 122-127, 1993

[6] Olensen et al. *A Minim Network for The Median Nerve: A Case Study on Loops* Applied Artificial Intelligence, 3:384-403, 1989

[7] Henrion, M. *Propagating Uncertainty in Bayesian Networks by Probabalistic Logic Sampling* Uncertainty in Artificial Intelligence 2, pp 149-163, NY, 1989

[8] Fung, R. and Chang, K-C *Weighing and Integrating Evidence for Stochastic Simulation in Bayesian Networks* Uncertainty in Artificial Intelligence 5, pp 209-219 NY, 1989

[9] Pearl, J. *Probabilistic Reasoning in Intelligent System s* Networks of Plausible Inference, San Mateo, CA, 1988

[10] Shachter, R. D. , and Peot, M. A. *Simulation Approaches to General Probabilistic Inference on Belief Networks* Uncertainty in Artificial Intelligence 5, pp 221-231 NY, 1989

[11] Russell, S. and Norvig P. *The Semantic of Belief Networks* Artificial Intelligence, A Modern Approach, pp 439, 1995